

Modeling and Analysis for Cache-Enabled Networks with Dynamic Traffic

Bin Xia, *Senior Member, IEEE*, Chenchen Yang, Tianyu Cao,

Abstract—Instead of assuming fully loaded cells in the analysis on cache-enabled networks with tools of stochastic geometry, we focus on the dynamic traffic in this letter. With modeling traffic dynamics of request arrivals and departures, probabilities of full-, free-, and modest-load cells in the large-scale cache-enabled network are elaborated based on the traffic queue state. Moreover, we propose to exploit the packets cached at cache-enabled users as side information to cancel the incoming interference. Then the packet loss rates for both the cache-enabled and cache-untenable users are investigated. The simulation results verify our analysis.

I. INTRODUCTION

Caching is a promising way to offload the wired backhaul traffic and the over-the-air traffic, especially for the highly-loaded network [1]–[3]. Poisson Point Process (PPP) of the stochastic geometry is a tractable tool to model and analyze wireless networks. The accuracy and tractability of PPP are elaborated in [4], [5], while the fully loaded cells (full queues at all times) are assumed. Based on this assumption, related studies have been conducted on the cache-enabled network. [6] focuses on the caching placement with stochastic geometry. The outage probability and the delivery rate are investigated with PPP for the backhaul-limited and cache-enabled network in [7]. [8] explores the storage-bandwidth tradeoffs with modeling locations of small cell base stations (BSs) as PPP. However, in previous studies, the assumption of fully loaded cells is too idealistic to fully capture the randomness of the stochastic network. Traffic dynamics of request arrivals and departures are studied in typical researches such as [3].

As BSs transmit wireless signals simultaneously, interference occurs and should be eliminated as much as possible. Cache-induced opportunistic cooperative multiple-input multiple-output (MIMO) is proposed for interference mitigation by caching parts of files at BSs in [9]. Caching split files at transmitters is considered in [10] to implement interference cancellation. [11] analyzes the benefits of caching and interference alignment at BSs in MIMO system with limited backhaul. Cache placement and interference management are elaborated in [12] where both transmitters and receivers are

cache-enabled. However, backhaul is needed for the inter-connection of cache-enabled transmitters to achieve cooperative transmission in previous studies. Interference cancellation at cache-enabled receivers to relieve backhaul requirement at the transmitter side remains to be studied.

In this letter, dynamic traffic is considered in the stochastic cache-enabled network, and the modeling method is proposed. Interference cancellation is performed at receivers using the cached packets. The main contributions are as follows:

- *Stochastic cache-enabled network with dynamic traffic:* Full-, free-, and modest-load cells are defined and the probabilities are analyzed based on queue states at BSs.
- *Interference cancellation with cached packets at receivers:* We propose the protocol that cache-enabled users exploit cached packets as side information to cancel incoming interference. The packet loss rates are analyzed.

II. SYSTEM MODEL

Consider a large-scale network, where BSs and users are spatially distributed based on two mutually independent PPPs Φ_b and Φ_u , with intensity φ_b and φ_u , respectively. Dynamic traffic is considered and packet (file) requests of each user are triggered based on the Poisson process with parameter λ [packets/second] [3], [13]. Users request packets from a given library $\mathcal{C} \triangleq \{c_1, c_2, \dots, c_C\}$ and all packets have the same size T [Mbits]¹. Each user randomly requests one packet in each time. Different packets are accessed by the user with different frequencies, which can be defined with the packet popularity distribution $\mathcal{F} \triangleq \{f_1, f_2, \dots, f_C\}$ for $f_i \in [0, 1]$ and $\sum_{i=1}^C f_i = 1$. Without loss of generality, we assume f_i is the popularity of packet c_i and $f_1 \geq f_2, \dots, \geq f_C$.

We focus on the downlink transmission, and the system is time-slotted with fixed duration τ [seconds]. Each BS transmits one packet in a slot with transmission power P . A user obtains packets from the BS providing the highest long-term average received signal strength (i.e., from the nearest BS). Requests are queued in the infinite buffers of the BS until they can be served. The service discipline is FIFO (first-in, first-out) [3]. Make the assumption that new request arrivals during a slot can not start to be served until the beginning of the next slot.

Consider a part ($0 < \alpha < 1$ proportion) of users are cache-enabled and each of them has a limited caching storage with size of $M \times T$ [Mbits]. When the network is off-peak, e.g., the nighttime, the M most popular packets are broadcasted via the BSs to and then pre-cached at the cache-enabled users. Denote

This work has been accepted by IEEE Communications Letters. Manuscript received June 02, 2016. The associate editor coordinating the review of this letter and approving it for publication was Alexey Vinel.

This work is supported in part by the National Key Science & Technology Specific Program (Grant No. 2016ZX03001015) and the Shenzhen-Hong Kong Innovative Technology Cooperation Funding (Grant No. SGLH20131009154139588).

The authors are with the Department of Electronic Engineering, Shanghai Jiao Tong University, Shanghai, China. Emails: {bxia, zhanchifeixiang, sherlockdoylecaoty}@sjtu.edu.cn. C. Yang is the corresponding author.

¹For ease of illustration, we assume that all packets have the same size. Results in the letter can be extended to the case of different packet sizes.

the set of the cached packets as $\mathcal{M} = \{c_1, c_2, \dots, c_M\}$. When the requested packet is available in the user's local caching set \mathcal{M} , the user obtains it immediately; otherwise, the BS transmits the requested packet to the user via the wireless downlink². Denote the cache hit ratio of the cache-enabled user as $\delta = \sum_{i=1}^M f_i$.

III. PROBABILITIES OF THE FREE-, FULL- AND MODEST-LOAD CELLS

Then BS cells are polygonal and form the Voronoi tessellation in \mathbb{R}^2 . The size of the BS cell is a random variable, which can be accurately predicted by the probability density function (PDF) $f_S(s) = \frac{(\varphi_b K)^K}{\Gamma(K)} s^{K-1} e^{-\varphi_b K s}$, $0 \leq s < \infty$. s denotes the cell size, $K = 3.575$ is a constant factor and $\Gamma(\cdot)$ is the gamma function. We then have the following lemma [14],

Lemma 1: The probability mass function (PMF) of the number of users in the coverage of a BS is

$$\mathbb{P}(N = n) = \frac{\varphi_u^n (K \varphi_b)^K \Gamma(K + n)}{(\varphi_u + K \varphi_b)^{K+n} \Gamma(n+1) \Gamma(K)}. \quad (1)$$

Proof: With the property of PPP and conditioned on the cell size, the number of users N in a BS cell is a Poisson random variable, which is generated by the conditional probability mass function (PMF), $\mathbb{P}(N = n | S = s) = (\varphi_u s)^n \exp\{-\varphi_u s\} (n!)^{-1}$, $n = 1, 2, \dots$. The PMF of the number of users in a BS cell is $\mathbb{P}(N = n) = \int_0^\infty \mathbb{P}(N = n | S = s) f_S(s) ds$. Then the proof is finished. ■

By noting that α of the users are cache-enabled and M packets have been cached into the cache-enabled users, the packet arrivals at a BS are Poisson process with parameter $\lambda_{b|n} = (1 - \alpha\delta)n\lambda$, conditioned on that the number of users in the coverage of a BS is $N = n$. Denote the number of requests in the queue at a BS after the k -th slot as q_k , that after the slot $(k+1)$ as q_{k+1} . The total number of request arrivals at a BS during the slot $(k+1)$ is denoted by a_{k+1} . So the a_{k+1} 's are independent identically distributed Poisson variables with parameter $\lambda_{b|n}$. Further we consider the scheme that a request will be deleted from the buffer no matter whether the requested packet has been transmitted completely at the end of the slot scheduled to it. We then have $q_{k+1} = [q_k - 1]^+ + a_{k+1}$. Here, $[x]^+ \triangleq \max\{x, 0\}$. In order for the BS to reach a stochastic equilibrium (to keep steady), the average number of requests arriving in a slot must be lower than the average number of packets that a BS can transmit out [3], [15]. It implies that the stochastic equilibrium is possible if and only if $\lambda_{b|n}\tau < 1$. It provides the critical condition (i.e., $(1 - \alpha \sum_{i=1}^M f_i)\lambda\tau n < 1$) for a BS cell to keep steady. Let $\bar{\lambda}$ denote $(1 - \alpha \sum_{i=1}^M f_i)\lambda$, i.e., $\bar{\lambda} \triangleq (1 - \alpha \sum_{i=1}^M f_i)\lambda$. We then have the following definition to clarify cells into three different types,

Full-load cell: Define the cell with the queue such that $\bar{\lambda}\tau n \geq 1$ as full-load cell. Since the BSs out of equilibrium are with infinitely-backlogged packets, we consider that they are transmitting packets at all times due to the saturation condition.

Free-load cell: Define the cell with the queue state such that i) $\bar{\lambda}\tau n < 1$, ii) the queue is empty, as free-load cell (i.e.,

there is no request waiting to be served in the queue). The BSs need not transmit packet in the free-load state.

Modest-load cell: Define the cell with non-empty queue such that $\bar{\lambda}\tau n < 1$ as modest-load cell. The BSs need transmit packet in the corresponding slot.

Therefore, based on the definition above, the probability that a randomly selected cell is a full-load cell is

$$p_t = \sum_{n=\lceil(\bar{\lambda}\tau)^{-1}\rceil}^{\infty} \mathbb{P}(N = n), \quad (2)$$

where the ceiling function $\lceil x \rceil$ is the smallest integer greater than or equal to x . It can be observed that p_t is jointly affected by the physical layer and information-centric parameters, e.g., $\varphi_u, \varphi_b, \lambda, \alpha, f_i$ and M . Moreover, p_t also means the fraction of the full-load cells in the two-dimensional plane.

To further get the probability of the free-load BS cell, we analyze the probability that the queue at the BS of stochastic equilibrium is empty. Conditioned on the number ($N = n$) of users in the coverage of a BS, define the conditional probability generating function (PGF) $Q(z|n)$ of the number of requests in the queue at the equilibrrious BS (i.e., $n < \lceil(\bar{\lambda}\tau)^{-1}\rceil$) as

$$Q(z|n) \triangleq \mathbb{E}[z^q|n], \quad \text{where, } q|n \triangleq \lim_{k \rightarrow \infty} q_k|n. \quad (3)$$

Here, $\mathbb{E}[x]$ is the expectation over variable x . Therefore,

Theorem 1: The probability of the free-load BS cells is

$$p_0 = \sum_{n=0}^{\lceil(\bar{\lambda}\tau)^{-1}\rceil-1} \mathbb{P}(N = n)(1 - \lambda_{b|n}\tau). \quad (4)$$

Proof: According to the definition in (3), we have

$$\begin{aligned} Q(z|n) &= \lim_{k \rightarrow \infty} \mathbb{E}[z^{q_{k+1}}|n] \stackrel{(a)}{=} \mathbb{E}[z^{a+[q-1]^+}|n] \\ &\stackrel{(b)}{=} \mathbb{E}_a[z^a|n] \mathbb{E}_q[z^{[q-1]^+}|n] \triangleq A(z|n)H(z|n). \end{aligned} \quad (5)$$

In Step (a), a is a random variable with the same distribution of a_{k+1} . Step (b) is obtained by noting that a_{k+1} and q_k are mutually independent random variables, so the same holds for variables a and q . Moreover, $H(z|n) = \mathbb{E}_q[z^{[q-1]^+}|n] = \sum_{i=0}^1 \mathbb{P}_{i|n} + \sum_{i=2}^{\infty} \mathbb{P}_{i|n} z^{i-1} = (1 - z^{-1})\mathbb{P}_{0|n} + Q(z|n)z^{-1}$, where $\mathbb{P}_{i|n} \triangleq \text{Prob}[q = i|n]$. Substituting it to (5), we have $Q(z|n) = A(z|n)\mathbb{P}_{0|n}(z - 1)[z - A(z|n)]^{-1}$. Substituting $Q(z|n)|_{z=1} = \sum_{i=0}^{\infty} \mathbb{P}_{i|n} z^i|_{z=1} = 1$ to $Q(z|n)$ and by noting that $A(z|n) = \exp\{-\lambda_{b|n}\tau(1 - z)\}$, we have $\mathbb{P}_{0|n} = 1 - \lambda_{b|n}\tau$. We then get (4) with $p_0 = \sum_{n=0}^{\lceil(\bar{\lambda}\tau)^{-1}\rceil-1} \mathbb{P}(N = n)\mathbb{P}_{0|n}$. ■

Moreover, based on (2) and (4), the probability of the modest-load BS cells is $p_m = 1 - p_0 - p_t$.

IV. PACKET LOSS RATE WITH AND WITHOUT INTERFERENCE CANCELLATION

Given the probabilities of BSs in different states, the actually active BSs in each slot are located as thinning PPP with parameter $\varphi_a = [p_t + (1 - p_t)(1 - p_0)]\varphi_b = [1 - p_0 + p_0 p_t]\varphi_b$. Without loss of generality, according to Slivnyak's theorem [4], [5], we assume that there is a typical user u_0 with or without caching ability at the origin of the Euclidean area.

²Extra cost (e.g., delay) for the BS to fetch packets from servers via wired backhaul is assumed to be neglected. It can be a promising topic for the future.

We first analyze the scenario where the typical user u_0 is cache-untenable. The received signal of u_0 is given by

$$y_0 = \sqrt{P}d_{0,0}^{-\frac{\beta}{2}}h_{0,0}x_0 + \sum_{k \in \Phi_a \odot b_0} \sqrt{P}d_{k,0}^{-\frac{\beta}{2}}h_{k,0}x_k + n_0, \quad (6)$$

where b_0 is the serving BS of user u_0 , $d_{0,0}$ is the distance between u_0 and b_0 . $\beta \geq 2$ denotes the path-loss exponent. We consider Rayleigh fading channel from b_0 to u_0 with average unit power, $|h_{0,0}|^2 \sim \exp(1)$. x_0 is the transmit signal from b_0 to u_0 with unit power. $d_{k,0}$ is the distance between interfering BS k and u_0 . Then interfering BSs are distributed according to PPP Φ_a with parameter φ_a outside the circle which is centered at the origin and with radius $d_{0,0}$ (denoted as $\Phi_a \odot b_0$). $|h_{k,0}|^2 \sim \exp(1)$ is the channel fading between interfering BS k and u_0 , x_k is the transmit signal from interfering BS k to user u_0 with average unit power, and $n_0 \sim \mathcal{CN}(0, \sigma^2)$ is the zero-mean additive white Gaussian noise (AWGN) with power σ^2 . In addition, we consider the channels remain constant in each slot. Then the signal-to-interference-plus-noise ratio (SINR) of the cache-untenable user u_0 is

$$\text{SINR}_u = \frac{P|h_{0,0}|^2 d_{0,0}^{-\beta}}{\sum_{k \in \Phi_a \odot b_0} P|h_{k,0}|^2 d_{k,0}^{-\beta} + \sigma^2} \triangleq \frac{P|h_{0,0}|^2 d_{0,0}^{-\beta}}{I_u + \sigma^2}. \quad (7)$$

I_u is the cumulative interference from the interfering BSs.

Next, we analyze the scenario where the typical user is cache-enabled. Note that α of users are cache-enabled. δ of their requested packets can be obtained immediately from local caching and yet $1 - \delta$ of the requested packets need to be obtained from the BSs. On the other hand, $1 - \alpha$ of users are cache-untenable and all of their requests need to be served by the BSs. δ of requested packets triggered by the cache-untenable users are the same with that in the caching set \mathcal{M} . That is, the requested packets of cache-untenable users can be divided into two subsets, one subset includes the same packets in $\mathcal{M} = \{c_1, c_2, \dots, c_M\}$, another subset includes the packets in the complementary set $\mathcal{C} \setminus \mathcal{M} = \{c_{M+1}, c_{M+2}, \dots, c_C\}$. In other words, on average, if \mathcal{P} requests are triggered by users in a slot, $\alpha\delta\mathcal{P}$ packets are obtained immediately from the local caching, and that need to be obtained from BSs are given by $[\alpha(1 - \delta) + (1 - \alpha)]\mathcal{P} = \{\alpha(1 - \delta) + [(1 - \alpha)(1 - \delta) + (1 - \alpha)\delta]\}\mathcal{P} = (1 - \delta)\mathcal{P} + (1 - \alpha)\delta\mathcal{P} \triangleq \mathcal{P}_{a1} + \mathcal{P}_{a2}$. The \mathcal{P}_{a2} and the \mathcal{P}_{a1} packets fall in the caching set \mathcal{M} and the complementary set $\mathcal{C} \setminus \mathcal{M}$, respectively. Then the BSs transmitting the two subsets of packets are respectively distributed according to thinning PPP Φ_{a2} with intensity $\frac{\mathcal{P}_{a2}}{\mathcal{P}_{a1} + \mathcal{P}_{a2}}\varphi_a$ and thinning PPP Φ_{a1} with intensity $\frac{\mathcal{P}_{a1}}{\mathcal{P}_{a1} + \mathcal{P}_{a2}}\varphi_a$. When the cache-enabled user requests an uncached packet, the received signal is given by

$$y_0 = \sqrt{P}d_{0,0}^{-\frac{\beta}{2}}h_{0,0}x_0 + \sum_{j \in \Phi_{a1} \odot b_0} \sqrt{P}d_{j,0}^{-\frac{\beta}{2}}h_{j,0}x_j + \sum_{k \in \Phi_{a2} \odot b_0} \sqrt{P}d_{k,0}^{-\frac{\beta}{2}}h_{k,0}x_k + n_0. \quad (8)$$

The interferences respectively come from the BSs distributed with Φ_{a1} and Φ_{a2} outside the circle which is centered at the origin and with radius $d_{0,0}$ (denoted as $\Phi_{a2} \odot b_0$ and $\Phi_{a1} \odot b_0$).

We consider the ideal case that the channel state information (CSI) knowledge is available at the cache-enabled users. We

further assume that, before interfering BSs in Φ_{a2} transmit packets that have been cached in the cache-enabled users, the cache-enabled users can be informed of the the packet indexes via extra interactive signals. Then the incoming interference from the BSs in Φ_{a2} can be canceled by cache-enabled users with the side information (i.e., cached packets in local caching). So the SINR of the typical user is

$$\text{SINR}_c = \frac{P|h_{0,0}|^2 d_{0,0}^{-\beta}}{\sum_{j \in \Phi_{a1} \odot b_0} P|h_{j,0}|^2 d_{j,0}^{-\beta} + \sigma^2} \triangleq \frac{P|h_{0,0}|^2 d_{0,0}^{-\beta}}{I_c + \sigma^2}. \quad (9)$$

I_c is the cumulative interference from the interfering BSs.

Definition 1: The packet loss rate \mathcal{P}_l is defined as the probability that a packet can not be transmitted completely at the end of the slot scheduled to it, i.e.,

$$\begin{aligned} \mathcal{P}_l &\triangleq \mathbb{E} \left[\mathbb{P} \left[\tau B \log_2(1 + \text{SINR}) < T \right] \right] \\ &= \mathbb{E} \left[\mathbb{P} \left(\text{SINR} < 2^{\frac{T}{\tau B}} - 1 \right) \right], \end{aligned} \quad (10)$$

where B is the bandwidth. Denote $\bar{T} \triangleq 2^{\frac{T}{\tau B}} - 1$ hereafter. The average is taken over both the spatial PPP and the channel fading. Firstly, we analyze the packet loss rate (PLR) for the typical user without caching ability. Denote r as the distance between user u_0 and its serving node. The PDF of r is $f_R(r) = 2\pi\varphi_b r \exp\{-\pi\varphi_b r^2\}$ [4]. We then have the following theorem,

Theorem 2: The PLR of the cache-untenable user is

$$\begin{aligned} \mathcal{P}_{l,u} &= 1 - \\ &2\pi\varphi_b \int_0^\infty r \exp\left\{-r^\beta P^{-1}\bar{T}\sigma^2 - \pi r^2 [\varphi_a \mathcal{Z}_1(\bar{T}) + \varphi_b]\right\} dr, \end{aligned} \quad (11)$$

where $\mathcal{Z}_1(\bar{T}) = \frac{2\bar{T}}{\beta-2} {}_2F_1[1, 1 - \frac{2}{\beta}; 2 - \frac{2}{\beta}; -\bar{T}]$, and ${}_2F_1[\cdot]$ denotes the Gauss hypergeometric function.

Proof: With the definition in (10), we have

$$\begin{aligned} \mathcal{P}_{l,u} &= 1 - \mathbb{E} \left[\mathbb{P}(\text{SINR}_u \geq \bar{T}) \right] \\ &= 1 - \int_0^\infty e^{-r^\beta P^{-1}\bar{T}\sigma^2} \mathcal{L}_{I_u}[r^\beta P^{-1}\bar{T}] f_R(r) dr. \end{aligned} \quad (12)$$

Here, the interference comes from all of the actually active BSs spatially distributed as PPP $\Phi_a \odot b_0$ with density φ_a . So the Laplace transform $\mathcal{L}_{I_u}[r^\beta P^{-1}\bar{T}]$ is

$$\begin{aligned} \mathcal{L}_{I_u}[r^\beta P^{-1}\bar{T}] &= \mathbb{E}_{I_u} \left[\exp(-r^\beta P^{-1}\bar{T} I_u) \right] \\ &= \mathbb{E}_{\Phi_a, \{|h_{k,0}|^2\}} \left[\exp\left(-r^\beta P^{-1}\bar{T} \sum_{k \in \Phi_a \odot b_0} P|h_{k,0}|^2 d_{k,0}^{-\beta}\right) \right] \\ &= \exp\left[-2\pi\varphi_a \int_r^\infty \left(1 - \frac{1}{1 + r^\beta \bar{T} v^{-\beta}}\right) v dv\right] \\ &= \exp\left[-\pi\varphi_a r^2 \mathcal{Z}_1(\bar{T})\right]. \end{aligned} \quad (13)$$

Accordingly, we have (11) and the proof is finished. ■

It can be observed that the packet loss rate is jointly decided by the physical layer parameter (e.g., BS density φ_b , user density φ_u , transmit power P , the slot duration τ and pass-loss β) and the information-centric parameter (e.g., the fraction of cache-enabled users α , caching ability M , packet popularity f_i , request rate λ and the packet size T).

Corollary 1: For the interference-limited network (i.e., $\sigma^2 \rightarrow 0$) and $\beta = 4$, the PLR is given by

$$\mathcal{P}_{l,u} = \frac{1}{\left[(1-p_0+p_0p_t) \sqrt{2^{\frac{T}{\tau B}} - 1} \arctan(\sqrt{2^{\frac{T}{\tau B}} - 1}) \right]^{-1} + 1}. \quad (14)$$

Similarly, for cache-enabled users, we have,

Theorem 3: When the cache-enabled user access the uncached packet from the BS, the PLR is

$$\mathcal{P}_{l,c} = 1 - 2\pi\varphi_b \int_0^\infty r \exp\left\{ -r^\beta P^{-1} \bar{T} \sigma^2 - \pi r^2 \right\} \times \left[\frac{\mathcal{P}_{a1}}{\mathcal{P}_{a1} + \mathcal{P}_{a2}} \varphi_a Z_1(\bar{T}) + \varphi_b \right] dr. \quad (15)$$

Proof: The derivation can be conducted with referring to the analysis of Theorem 2, and the proof is omitted here. ■

Corollary 2: When $\sigma^2 \rightarrow 0$ and $\beta = 4$, the PLR is,

$$\mathcal{P}_{l,c} = \frac{1}{\left[\frac{(1-p_0+p_0p_t)}{(1-\alpha\delta)(1-\delta)^{-1}} \sqrt{2^{\frac{T}{\tau B}} - 1} \arctan(\sqrt{2^{\frac{T}{\tau B}} - 1}) \right]^{-1} + 1}. \quad (16)$$

Compare (11) with (15), it can be observed that the packet loss rate of the cache-enabled user is lower, owing to the interference cancellation with the cached packets.

V. NUMERICAL RESULTS AND DISCUSSIONS

The simulation results are obtained with Monte Carlo methods in a square area of $10^4\text{m} \times 10^4\text{m}$, where the nodes are scattered based on PPPs with intensity of $\{\varphi_u, \varphi_b\} = \{\frac{400}{\pi 500^2}, \frac{4}{\pi 500^2}\}$ nodes/ m^2 . The transmit power is $P = 43$ dBm and 20 MHz bandwidth are shared among different BSs. We set the path-loss $\beta = 4$, the packet arrival rate $\lambda = 0.025$ packets/second, the fraction of cache-enabled users $\alpha = 0.25$ and the slot duration $\tau = 0.5$ second. We consider the packet popularity follows Zipf distribution $f_i = \frac{1/i^\gamma}{\sum_{j=1}^C 1/j^\gamma}$, where $\gamma = 0.8$ and the total number of packets $C = 200$. The size of each packet $T = 10$ Mbits and the caching ability $M = 10$. Consider the network where caching is not available at users as the baseline, i.e., the fraction of cache-enabled users $\alpha = 0$.

The probabilities of free-, full- and modest-load BS cells are illustrated in Fig. 1. The analysis and simulation results are consistent well. Compared with the baseline, the probabilities of full-load cells decreases while that of the free- and modest-load cells increases due to the traffic offloading via caching.

In Fig. 2, it can be observed the packet loss rate of cache-enabled users is far lower than that of the cache-untenable users and the baseline, owing to the interference cancellation gain with caching. When $M = 5$ (15), the average loss rate of the cache-enabled network decreases by 9.80% (15.46%) compared to that of the baseline. Excluding free-load BSs out of the interfering nodes, our analysis result for the packet loss rate is lower than that of previous studies regarding all cells are fully loaded. Our work is a supplement to previous studies.

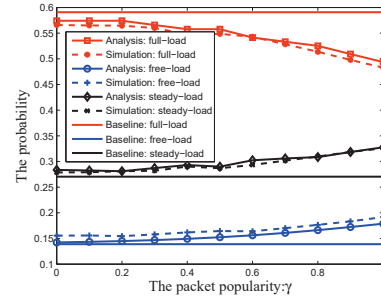


Fig. 1. The probabilities of BSs in full-, modest- and free-load state for the caching network and the baseline. Note that previous studies consider all cells are active in full-load state, i.e., the probabilities of full-, modest- and free-load BSs are $\{100\%, 0\%, 0\%\}$.

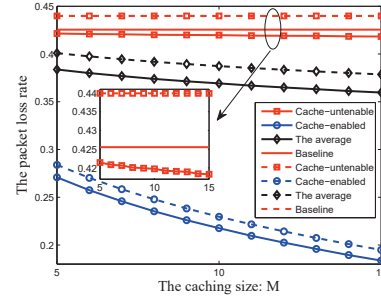


Fig. 2. The packet loss rate with the change of caching size: solid lines are results of our analysis, dashed lines are that of previous studies regarding all cells are fully loaded. Note that for previous studies, the packet loss rate of the baseline overlaps with that of cache-untenable users [4], [5].

REFERENCES

- [1] N. Golrezaei, A. Molisch, A. Dimakis, and G. Caire, "Femtocaching and device-to-device collaboration: A new architecture for wireless video distribution," *IEEE Commun. Mag.*, vol. 51, no. 4, pp. 142–149, Apr. 2013.
- [2] E. Baştuğ, M. Bennis, and M. Debbah, "Living on the Edge: The Role of Proactive Caching in 5G Wireless Networks," *IEEE Commun. Mag.*, vol. 52, no. 8, pp. 82–89, Aug. 2014.
- [3] A. Vinel, Q. Ni, D. Staehle, and A. Turlikov, "Capacity analysis of reservation-based random access for broadband wireless access networks," *IEEE J. Sel. Areas Commun.*, vol. 27, no. 2, pp. 172–181, Feb. 2009.
- [4] H.-S. Jo, Y. J. Sang, P. Xia, and J. Andrews, "Heterogeneous Cellular Networks with Flexible Cell Association: A Comprehensive Downlink SINR Analysis," *IEEE Trans. Wireless Commun.*, vol. 11, no. 10, pp. 3484–3495, Oct. 2012.
- [5] H. ElSawy, E. Hossain, and M. Haenggi, "Stochastic Geometry for Modeling, Analysis, and Design of Multi-Tier and Cognitive Cellular Wireless Networks: A Survey," *IEEE Commun. Surveys and Tutorials*, vol. 15, no. 3, pp. 996–1019, June 2013.
- [6] B. Błaszczyszyn and A. Giovanidis, "Optimal Geographic Caching in Cellular Networks," in *Proc. ICC*, June 2015.
- [7] E. Baştuğ, M. Bennis, K. Marios, and M. Debbah, "Cache-enabled small cell networks: modeling and tradeoffs," *EURASIP Journal on Wireless Commun. and Networking*, Feb. 2015.
- [8] S. T. ul Hassan, M. Bennis, P. H. J. Nardelli, and M. Latva-aho, "Caching in Wireless Small Cell Networks: A Storage-Bandwidth Tradeoff," *IEEE Commun. Letters*, vol. 20, no. 6, pp. 1175–1178, Mar. 2016.
- [9] A. Liu and V. K. N. Lau, "Mixed-Timescale Precoding and Cache Control in Cached MIMO Interference Network," *IEEE Trans. Signal Process.*, vol. 61, no. 24, pp. 6320–6332, Dec. 2013.
- [10] M. A. Maddah-Ali and U. Niesen, "Cache-Aided Interference Channels," in *ISIT*, June 2015.
- [11] M. Deghel, E. Baştuğ, M. Assaad, and M. Debbah, "On the benefits of edge caching for MIMO interference alignment," in *Int'l Workshop on Signal Proc. Advances in Wireless Commun. (SPAWC)*, June 2015.

- [12] N. Naderializadeh, M. A. Maddah-Ali, and A. S. Avestimehr, "Fundamental Limits of Cache-Aided Interference Management," in *arXiv:1602.04207 [cs.IT]*, <http://arxiv.org/abs/1602.04207>, 2016.
- [13] M. Karray and M. Jovanovic, "A Queueing Theoretic Approach to the Dimensioning of Wireless Cellular Networks Serving Variable-Bit-Rate Calls," *IEEE Trans. Veh. Technol.*, vol. 62, no. 6, pp. 2713–2723, July 2013.
- [14] H. ElSawy and E. Hossain, "Two-Tier HetNets with Cognitive Femto-cells: Downlink Performance Modeling and Analysis in a Multichannel Environment," *IEEE Trans. Mobile Computing*, vol. 13, no. 3, pp. 649–663, Mar. 2014.
- [15] H. Bruneel, "A General Model for The Behaviour of Infinite Buffers with Periodic Service Opportunities," *European Journal of Operational Research*, vol. 16, no. 1, pp. 98–106, 1984.